#22 Data-driven attribution - the past, present and future (with Tom Woods)

[00:00:17] **Dan:** Hello and thanks for joining us in The Measure Pod, a podcast for analytics enthusiasts. I'm Dan, I'm an Analytics Consultant at Measurelab. I'm back this week without Dara, he's still on holiday, but hopefully will be back with us next week. In the meantime, I've brought on an old friend and colleague of mine from many, many moons ago, Tom Woods, to talk to us about all things attribution. So Tom, it's good to have you on the pod mate.

[00:00:41] **Tom:** Dan mate, thanks for having me. I'm very happy to be here.

[00:00:44] **Dan:** My pleasure, what I always like to ask people is how you ended up in this world of analytics. I find it a fascinating question to ask because it's not a traditional career path. So Tom, how did you end up in analytics?

[00:00:56] **Tom:** Well, everybody else wanted to be astronauts and I want it to be an analytics bod. Ah no, I, studied marketing and advertising. However, the whole creative side of things was never really my forte. And it was always the why's of advertising that I've found interesting and hypothesizing why people buy things and the different routes that people take. So I studied that in uni, and then after that did a couple of odds and sods jobs, as everybody does after uni, and then applied for an internship at a company called DC Storm. Which was where mine and your stars crossed back in like 2014, we worked together for a number of years on a number of attribution clients. And then after that jumped ship when the company disbanded and then I now work at an agency called Jellyfish doing lots of data planning, lots of interesting data activation pieces for paid media clients.

[00:01:53] **Dan:** It's one of those industries that grabs hold and never lets go, isn't it. It's a sticky industry. It grabs that person that's into problem solving, puzzle solving, challenges you have to go figure out for yourself.

[00:02:03] **Tom:** That's it, it's the data manipulation side of things as well, I find quite cathartic. I've always found that element of what we do really interesting.

[00:02:12] **Dan:** Great. Well, it's been no secret that I've wanted to chat to you on this show for a while, Tom. And I think it'll be a really good conversation because both of our careers started in the same space, which was attribution modeling. Everything was talking about how attribution is the answer to all marketing problems. That was 8, 10 years ago and the industry has changed, the world has changed, the technology has changed, we have also changed. And I thought this would be a really good time to talk about attribution, it seems to be the word of the month again as it was back years ago, where it seemed to be the buzzword of the year. Whereas now it's the buzzword of the month, things go a bit quicker now. And that's just because Google have released their data-driven attribution as a default and recommend property setting now. So anyone with GA4 right now, you can go into your property settings and you can have to change the default attribution. But I think starting right at the beginning Tom, when we were talking attribution, it was a different world. It was something that we could literally track everything about everybody that came to a website or even apps as well at that point. And then we could do attribution modeling across everything from ad impressions, to ad clicks, to on-site activity, to purchases, even offline data, we even ingested some TV ads data. There was no concept of, well maybe there was, whether this is okay to collect. Can we join this personally identifiable information and join that across multiple different data sources, store it in a third party that almost wasn't part of the conversation back then. But it enabled us to do some cool stuff, but I think since then, the thoughts around attribution has remained the same, but everything else has changed. So I just really want to get your thoughts on, how looking back in hindsight, how relevant was attribution in the grand scheme of things. And today, has that changed?

[00:03:58] **Tom:** It's a really interesting point. I think it's always been an ongoing methodology of measurement that people always rely on. Even if it's last click measurement, it's still ultimately tracking certain actions and then trying to figure out what to do off the back of it. It's interesting how, before GDPR and everything, it was like the wild wild west in that it was like a free fall. You could mash multiple sets of data and journeys together. Lot of cross device tracking that is quite hard to do now. It's interesting how marketing measurement was almost come full circle now back to direct marketing methodologies. Where whereas before we were able to get super granular and into the detail, which before attribution, and before cookies, was never really possible. Everything was done on econometric modeling and incrementality studies to try and figure out what is and isn't working. And now, because of the ultimate demise and future death of the cookie, everything's moving more back towards that direct marketing principles of incrementality studying alongside attribution. Attribution is still a core part of media measurement, and looking beyond the last click is still super important. However, because of the way things are tracked now and rightly so, putting the consumer at the heart of privacy policies and that sort of thing, it means that the data becomes a lot more disparate. It's now much more shorter, we're getting like a couple of touch points before somebody converts rather than being able to cast that wider net. So it's definitely different. However, there's still ultimately benefits to using attribution within measurement and not just relying on that last click.

[00:05:41] **Dan:** Is there something in that you mentioned around the view of a customer journey is now shorter. So if you are a retailer that has a very short sales cycle with your products, take like an Amazon where I want to buy something, then I'll just purchase it. It almost makes very little difference because when we think of sales cycles being shorter, it's unaffected by browser policy changes as well. Things like ITP and ETP that are setting expirees cookies a lot shorter than they have been set natively by the JavaScript code. But then if you're looking at a company that has a very long sales cycle, take a holiday or a cruise company with a very high ticket price, and if they had a lot of budget allocation based on this multi-touch attribution approach, they're probably going to be more effected. So it's going to be one of these things that for the people or for the businesses that are affected, that they're painfully aware that's already happened and they've already had to change things. But for the people that may not have realized that this has happened, it's probably not going to be that much of an issue anyway, and they're continuing as they are. You did mention one thing which I want to touch on as well, is the whole buzz sentence I'll call it, which is the death of the cookie, right? The cookie we're talking about is the third party cookies, first party cookies, aren't going anywhere. They have to exist the things to work. However, the third party cookies, what we're talking about, and it's pretty much only Chrome that have got this left and they're removing that in 2023. So there is a slow demise, but I would say it's on life support as it is. There's a there's a year or so of it ticking over. But when it's not like something crazy is going to happen at that point in time, considering most other browsers have already deprecated the third party cookie or blocked entirely.

[00:07:18] **Tom:** Yeah definitely, that's certainly a really important point to make. So you mentioned Amazon, it's funny how all of these different platforms now the walled gardens are getting taller and taller. Now Amazon trying to get advertisers to purchase directly through Amazon, and they have their own attribution solution as well, where you can look at the journey of Amazon DSP, so like display down to their own search ads as well. And you also then got Facebook as well, who are doing their own thing and have their own challenges with measurement. You've then got Google whereby the whole browser and logged in state is going to be super key for them to be able to make sure that that ultimately their measurement for their advertising is going to be future-proofed.

[00:08:00] **Dan:** Well, that's the Google Signals feature right, that you actually can layer into Google Analytics and is a key concept of GA4. Because from Universal, they use signals and they pulled in things like demographics and interest data, which was nice, but it's not super essential. What GA4 started to do is actually use that as a identity resolution. So you can actually use this walled garden, this black box of Google to do some identity resolution, which then ties into user journeys and attribution modeling. The way I see this, or the way that it's going is that Google Analytics is becoming partially or more and more as time goes on behind that walled garden. And I think this whole idea of a walled garden is something that if we're not super familiar with it now, we're going to have to be very quickly because it's creeping up on us fast. And this is this concept of, or at least it was called GAFA, which is the Google, Amazon, Facebook and Apple ecosystems, they're the big four players. I suppose we can call it GAMA now, with a rebrand to Meta for Facebook. Each of those operate independently, do their own modeling, attribution, retargeting, audience profiling. What we're going to have is a very limited version of that through consented first party data. So when we talk about first party data being key, it's saying, I've asked the customer can I track you and use this data. They have said, yes, and I create this first party ecosystem. Which would be, I suppose we can label that as a CDP a customer data platform, or you can call it just a database if you want, or a CRM on steroids, but it's about taking that data and sharing those out to your ad platforms for retargeting. But the whole concept of retargeting an unknown user that visits your website, that is going or gone. It's already gone. That's already something you can't do. And if you are doing that currently, then that's something we need to take a closer look at. But I think as you said Tom, which is absolutely bang on, the way I think about it too, is that we're going back in time in a way of marketing. We're going back to this contextual marketing. Like Meta have a way of saying I've got 25 to 35 year olds that have shown an interest in football, do you want to show an ad to them? And they're definitely not users that you can track and see. Meta, in this example, are letting you into their walled garden and said, spend some money with us and we can reach these people. And you've got no way of really validating that. And that's the hard thing, I suppose. I don't know if you see this as well Tom, but this is the hard thing is the change. It's hard to adapt from a place of having everything, going to not have everything and trying to keep things, keep things, going, keep things the same.

[00:10:23] **Tom:** That's it, it's a super step change, especially for a lot of marketing managers who this is what they've learn over time. So it is very much a real step change. And that move towards contextual buying something that yeah, if advertisers aren't starting to think about it, then they really should do. Yeah, it poses unique and interesting challenges in our industry. It's a fast moving industry and if there's any industry that's best place to figure it out with new data, it's people like us.

[00:10:51] **Dan:** Yeah, too right. Well enough about the glory days or reliving the past. I've got a question around how have you been using, or I suppose the first question is, have you been using things like algorithmic or data-driven attribution within the world of media buying in your role at Jellyfish? How is attribution being used for these big advertisers?

[00:11:13] **Tom:** It plays a part dependent very much on their tech setup. Predominantly it's Google, and then from there, we then work out what tools they have within the GMP that we use for attribution. So typically if it's a bigger paid media client, they will use Campaign Manager, SA360 and then we'll use the attribution modeling that's available within Campaign Manager, which is floodlight based, then gives you richer access into stuff like Ads Data Hub as well. But on the ground, we use the Google specific data-driven attribution modeling as our recommended source of truth across all of the different channels. It's black box in that we're not able to entirely validate its methodology, however we're able to validate it through stuff like conversion path analysis. But what it does do, is it takes out the biases that can happen within attribution modeling. So the rules based models such as last click, but also linear and first touch. The challenge inherently with those is that a marketing manager can ultimately pick how they want their marketing to be measured, whereas with algorithmic modeling, there is that black box element. However, if you use it and compare it against baseline sales, actually plot it alongside first party customer database sales. If you're doing the right things within paid media using a data-driven attribution, you should hopefully then see the increase in the base alongside of the sales that the other internal stakeholders look at like finance and procurement and that sort of thing.

[00:12:56] **Dan:** So when it comes to using these attribution models, I mean, we've got the concept of data-driven attribution or DDA as we'll see it plastered over all of these Google products now. They have rolled out. They are the default and the recommended from Google's perspective and agencies such as yourselves. I suppose, between the two of us, we could argue the toss between every different attribution model and probably convince someone to pick a first click or last click or linear, or data-driven. So when it comes to using these tools, let's stay within the Google walled garden, the Google ecosystem. We've got data-driven attribution in Google Ads. We've got data-driven attribution in Google Analytics now, we've got data-driven attribution in Campaign Manager. We've got data-driven attribution everywhere. Each one is different, but it's also called data-driven attribution. This is just an open question, but where do you see all of this going, at some point they have to align. Where do you see the Google stack going around attribution and this concept of data-driven? Are they going to keep them all quite separated as they have to date with their own applications, or is there a view of this where something like Google Analytics 4 come in, be the central hub for all this, and make something like Campaign Manager redundant. I don't know. Sorry. I'm maybe putting words in your mouth there, but this is where my heads are thinking about what is the future of all these stacks.

[00:14:14] **Tom:** That's an interesting point. Definitely crystal balling, I think you're right in the latter, in that it seems like GA4 is being set up to be that hub for attribution and for utilizing that first party data to then be able to activate in your advertising. I think it's quite telling that the module on the left-hand side of GA4 is called advertising. And even in naming terms, that's a bit of a step change for Google moving away from GMP and marketing to being like, these are things for you to use and activate specifically within advertising and specifically within Google advertising. And the new model in GA4 seems to be a bit more comprehensive than the GA360 one. My understanding is it's like the last 50 touch points versus the 4 you have in GA360 at the moment. Again, demonstrates that they're looking to make it more comprehensive. So almost in the middle between what is the GA360 one at the moment, and then the one within Campaign Manager where I think it's like 200 touch points or something like that. It feels like this is like a meeting at the middle almost, and the whole floodlight based ecosystem will be more important perhaps for specific custom modeling either within Ads Data Hub or the flock is that flock?

[00:15:40] **Dan:** Yeah, the, the federated learning of cohorts.

[00:15:44] **Tom:** Cohort based stuff, which again is going way back to the direct marketing principles so yeah, it definitely feels that GA4 is going to be like the one-stop shop, which is quite exciting, it's quite exciting. And also the goal based bidding as well. So, whereas at the moment within like SA360 or like the other paid media tools within the Google Marketing Platform, everything's based off a floodlights. You can do like a bit of goal based bidding I think, if you hook up GA360 to SA360. But the fact that you'd be able to do goal-based bidding directly into their platforms.s Again, just feels like everything was becoming a lot more Google Analytics centric.

[00:16:28] **Dan:** Yeah, and the other step change which you kinda touched on there is that everything's connected to GA360. The GA360 license opens up the DV360, SA360, Campaign Manager 360 connectors. Whereas in GA4, they're all going to be available for free on the free tier. So in a way, there's this blocker, this paywall that we have within GA360 on things like exporting your first party data out into BigQuery, connecting your conversions to not just Google Ads, but the rest of the GMP. Search Ads, DV, to then use for optimization taking away that barrier. I think when these features are rolled out, when these connectors are rolled out, when GA4 becomes the primary analytics tool for most advertisers, I think that we're going to see a huge adoption of things like DV360, SA360. I suppose even crystal balling that, there's a future where, why do you have all these different platforms? Why not just have Google Ads have everything connect that to your one analytics platform. So you've got one ad buying platform, one analytics platform, and maybe even roll those together, who knows. The thing is they could do whatever they want to with it right. We're actually still working with quite legacy tools. You know, you still see some of the Dart Search or DoubleClick redirects on some of the search ads in the wild. I still see those and, you know, w.

[00:17:46] **Tom:** guess.

[00:17:47] **Dan:** That's it. Yeah, I find it interesting now. So I've, I'm, I'm almost lost with all the different acronyms. It feels like we're going through that with Firebase. So Google bought Firebase a number of years ago. They've rolled that out for web, so that's become the new GA4. They've found this better technology, they've overtaken the old urchin analytics that they bought back in the nineties and they're replacing it. And I think this opens doors, right. And who knows where their roadmap takes them with all the other ad buying tools. One of the most interesting conversations Tom, I've had with Dara on this podcast was around the fuzziness of data. Data's becoming un-validatable, if that's even a word, because within Google Analytics 4, we can only track consented users. So if someone has to opt into the analytics cookies on the website for it to get into GA to begin with. And then we can track all sorts different things, we can export that to BigQuery, we can do all the clever, cool stuff there. But what Google has started to do is they've all introduced data driven attribution modeling, this black box algorithm. But on top of that, it's laid in something called conversion modeling, which it did a couple of months ago, and soon they're releasing behavioral modeling. And what this means is for all of those unconsented users, they're going to start modeling out the difference. So you'll still continue to report on quote unquote all users, but you actually haven't tracked them all. So you've basically got this model, based on model, based on a model. There's no real point there, sorry Tom. It was just a conversation around data-driven attribution is one model that's being applied on that. I'm struggling to justify in my own head right now, let alone to clients, to invest your money based on what this is saying. It's becoming hard, right? It's becoming hard to know where the human intervenes or how much control we can have of this.

[00:19:29] **Tom:** Yeah, it's a mess, it's a mess Dan. It's like if anyone listens is a fan of football, there's this model called XG, which is expected goals versus the goals that actually happen in a game. So it's sort of what should happen based off of things that have happened before. And basically you're just moving from absolutes to an indicative view of things. And I think that indicative view is important as a term, because it KPIs or key performance indicators, they're indicators of performance. Attribution before, was reporting very much in absolutes. Whereas now everything is based off of reasoned assumptions using the information that you have available, and ultimately attribution plays a part of that. But then there are other types like running incrementality studies for channels is a key thing. Before, you could just report a number and it was agreed on, you have to overlay the assumptions and I hypotheses to things on how things are tracked. So it's, I dunno, that was a bit of a ramble, but I feel as there feels like, it feels like ultimately there isn't going to be a real perfect source of truth anymore. And it's just down to marketeers and advertisers to agree on a methodology that they're happy with and ultimately what the rest of the business are happy with, and just using that.

[00:20:52] **Dan:** Yeah, I couldn't agree more. It's just about broadening your analysis horizons, or analytical horizons I think more than anything. We've been so used to this idea of multi-touch attribution being the only source of truth. Whereas what we're going to now is, again, something that harks back to methodologies in the past, is this idea of media mix modeling, or econometric modeling, where it's analysis of aggregate data. Daily data that we can see and identify correlation and trends to then infer some effect. And you also mentioned the incrementality studies. I think having this concept or this culture of always testing, always understanding what's the uplift of this so that we can look at, not just the multichannel attribution ROI, but the, the uplifted ROI. You have to broaden the horizons when it comes to the analysis, you have to incorporate different methodologies, understanding where the black boxes have value where these walled gardens have value and how to make them work for you rather than you continuously trying to feed this data into the walled garden.

[00:21:55] **Tom:** That's a good point now on attribution piece that if you're only looking at attribution and what's trackable within the ecosystem within that model in isolation. It's inherently lower funnel. So it's all going to be lower funnel, relatively direct response activity. And if you're basing decisions purely off of that, you're only just capturing demand that's already there, rather than actually going out, reaching new people through contextual buying, and through brand awareness activity and driving brand association and brand demand. The ultimately that then does get captured within those activation channels that the attribution lends itself to. You then have the next step up, which is running this always on incrementality testing. So that in Facebook, if you have Facebook Conversion API, or within different platforms, you can then validate the effectiveness of those impression buys that you don't get within the Google shop attribution. And then Dan you raised a really good point on MMM econometric modeling, we're seeing that shift at Jellyfish now. You know, advertisers are talking to us more and more about it. And the great thing about it is, there isn't that joining and stitching and that messiness that the data, it's all aggregate information and using the incrementality pieces and the brand lift pieces and the attribution pieces to validate the investment that's going through the pipes. Ultimately, if something's working or wash itself within that type of econometric modeling. So it's using the short term granular indicator and methodologies like attribution to justify the investment going through the pipes, and then you use big, broader methodologies to make the bigger quarterly, more annual budgeting decision, rather than the the short term shifts that we sometimes see within media buying.

[00:23:58] **Dan:** Exactly, that real time optimization cannot be through media mixed modeling or econometric models because that is a four year set of data that you're analyzing for trends. You can't be real time bidding per minute per hour, you'll need some other way of doing that. But yeah, I couldn't agree more. All right Tom, thank you so much for that chat, hopefully next time we won't have the mics, but we'll have a beer in our hands and you can give me the lowdown on what's happening with Ads Data Hub and Campaign Manager again.

[00:24:25] **Tom:** Yeah.

[00:24:26] **Dan:** Tom, the way we like to end the show is just to wind down, get rid of analytics from the brain, let's not talk about anything to do with the big G anymore. And just ask you what you've been doing to one down, what have you been in to switch off from analytics in the last couple of weeks?

[00:24:39] **Tom:** Tell you what, I've been digging out loads of old cookbooks and just getting back into making f un midweek food to break up the week rather than just relying on the old spag bol bits and bobs. That's been quite fun, like mining, lots of old cookbooks. In terms of watching stuff, I know George in the last podcast mentioned The Witcher, but me and my partner got absolutely hooked on it over the last week or so. Isn't really normally my bag, but I tell you what, to really switch off from numbers and go somewhere else, it was quite fun.

[00:25:10] **Dan:** Yeah, amazing. And as a Witcher fan, I'll ask you the same question I asked you, which have you played the games or read the books?

[00:25:17] **Tom:** Nah, man. I'm a pedestrian, I'm very much a pedestrian rather than a hardcore fan.

[00:25:24] **Dan:** Amazing. Well look, getting into that world is awesome, whatever your medium of choice is. But if you are a gamer, I highly recommend The Witcher 3.

[00:25:31] **Tom:** Dan, what are you been doing to switch off from the world of data analytics?

[00:25:36] **Dan:** Well, it's something that I've mentioned probably many times in this podcast before, but it has to be skateboarding. I've just been out skating all the time, bashing and injuring and bruising myself like crazy. Definitely feeling my age when it comes to this.

So that's it for this week, you can find out more about us and the podcast over at measurelab.co.uk, and you can get in touch with us directly using the email podcast@measurelab.co.uk, or find myself or Measurelab on LinkedIn and given us a message there. You can suggest a topic for us to talk about or suggest a guest to ask a couple on the show or even come on the show yourself, just drop us an email or message me on LinkedIn, and we will get that arranged. So join us next time. I'm Dan, I've been joined by Tom. So it's bye from me.

[00:26:21] **Tom:** And bye from me.

[00:26:22] **Dan:** See you next time.